IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON ARTIFICIAL INTELLIGENCE TECHNOLOGIES FOR ELECTRIC POWER SYSTEMS

Received September 29, 2019, accepted October 21, 2019, date of publication October 24, 2019, date of current version November 6, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2949480

Deep Reinforcement Learning-Based Tie-Line Power Adjustment Method for Power System Operation State Calculation

HUATING XU[1], ZHIHONG YU[1], QINGPING ZHENG[2], JINXIU HOU[1], YAWEI WEI[1], (Member, IEEE), AND ZHIJIAN ZHANG[3]

[1]China Electric Power Research Institute, Beijing 100192, China
[2]School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China
[3]State Grid Beijing Electric Power Dispatching and Control Center, Beijing 100031, China

Corresponding author: Zhihong Yu (zhhyu@epri.sgcc.com.cn)

This work was supported by the Adaptive Generation and Interactive Adjustment Technology of Power Grid Operation State Based on Online Data under Grant 5442XT190011.

ABSTRACT Operation state calculation (OSC) provides safe operating boundaries for power systems. The operators rely on the software-aid OSC results to dispatch the generators for grid control. Currently, the OSC workload has increased dramatically, as the power grid structure expands rapidly to mitigate renewable source integration. However, the OSC is processed with a lot of manual interventions in most dispatching centers, which makes the OSC error-prone and personnel-experience oriented. Therefore, it is crucial to upgrade the current OSC in an automatic mode for efficiency and quality improvements. An essential process in the OSC is the tie-line power (TP) adjustment. In this paper, a new TP adjustment method is proposed using an adaptive mapping strategy and a Markov Decision Process (MDP) formulation. Then, a model-free deep reinforcement learning (DRL) algorithm is proposed to solve the formulated MDP and learn an optimal adjustment strategy. The improvement techniques of ''stepwise training'' and ''prioritized target replay'' are included to decompose the large-scale complex problems and improve the training efficiency. Finally, five experiments are conducted on the IEEE 39-bus system and an actual 2725-bus power grid of China for the effectiveness demonstration.

INDEX TERMS Operation state calculation, tie-line power adjustment, deep reinforcement learning, stepwise training, prioritized target replay.

I. INTRODUCTION

Operators rely on operation state calculation (OSC) which provides the grids' safe operating boundaries to estimate the security level of power systems. In recent years, due to the development of the social economy, power consumption and the access of renewable energy have continuously set new records [1]–[3]. As a result, the grid structure expands markedly, making the number of typical operation modes (TOMs) and key transmission sections (KTSs) increase dramatically [4]–[7] as well. Therefore, it becomes a challenge to complete the OSC today.

At present, the OSC of large-scale power grids is still processed with a lot of manual interventions, and the calculation process can be divided into three stages: i. Forecast the load in the planned future; ii. Formulate TOMs based on the load forecasting results; iii. For each TOM, calculate the transfer capability limit (TCL) of each concerned KTS. In practice, the stage iii is mainly achieved by adjusting the tie-line power (TP) to different values manually and executing transient simulation under preset faults to find the safe operating boundaries. Due to the increasing amount of TOMs and KTSs, the OSC has become tedious, arduous, and repetitive requiring continuously updated operating experience in power grids' operation states. Therefore, it is crucial to develop an algorithm that can automatically complete the OSC.

In recent years, numerous researches have focused on the OSC from different perspectives. For TOM formulation, Y. Zhang *et al.* [8] proposed a method for the integration of multi-source data and auto-adjustment of power flow to formulate the typical operation modes. H. Wang *et al.* [9]

The associate editor coordinating the review of this manuscript and approving it for publication was Mingjian Cui.

proposed a concept of approximate power flow (APF) to deal with the convergence problem of power flow calculation, which improves the efficiency of TOM formulation significantly. Ren and Zhang [10] devised a generalized microgrid power flow with good convergence for analyzing the islanded microgrid. Taking into account the impact of renewable energy uncertainties and load forecasts, Reddy and Momoh [11] proposed an optimum day-ahead scheduling strategy for a hybrid power system to minimize both day-ahead and real-time adjustment costs. Lee *et al.* [12] introduced a bus-dependent participation factor based on generation-load incremental cost and proposed a new generation adjustment approach to operate the power system economically. For the TCL calculation, Wang and Gao [13] proposed a new method of analyzing the available transfer capability of AC-DC power systems and the adjustment of HVDC control which can provide a reference for power system dispatch and operation. Liu *et al.* [14] used nonparametric analytics to estimate the TCL of a power system based on online measurement. Xu and Miao [15] presented the multi-area TCL calculation method based on improved Ward-PV equivalents. Although the approaches above contribute a lot to the OSC from different perspectives, it is hard to take into consideration all the constraints of practical scenarios, especially in large-scale AC-DC hybrid power systems. There has not been a method that can complete the OSC without manual interventions.

Recently, model-free methods that do not depend on system model information have achieved great success in solving complex decision-making problems [16], [17]. These achievements have inspired the development of model-free methods in power systems [21]. For the power system operation, R. Yousefian *et al.* [18] proposed a Wide Area Control design based on reinforcement learning (RL) and neural network to enhance the transient stability of power systems integrated with doubly fed induction generators. Zhang *et al.* [19] introduced a load shedding scheme against voltage instability based on deep reinforcement learning (DRL) using spatial and temporal information. Yan an Yu [20] used DRL in a continuous action domain to minimize the frequency deviation with stronger adaptability and quicker response speed. For the electricity market, Wan *et al.* [21] formulated the real-time EV charging scheduling problem as a Markov Decision Process (MDP) and determined the optimal strategy based on a representation network and a Q network. Ruelens *et al.* [22] applied the RL algorithm to control an electric water heater with lower energy consumption in practice. The reinforcement learning was also applied in energy and load management [23]–[25]. More applications in power systems are introduced in [26]. Overall, the DRL based model-free methods have achieved success in the complex decision-making problems of power systems. Similarly, the OSC involves lots of decision-makings as well. Nevertheless, to the best of our knowledge, few applications of DRL in the OSC have been reported in the literature.

This paper focuses on upgrading the transfer capability limit (TCL) calculation process with a model-free method. From stage three of the OSC, it is easy to know that the key to calculate the TCL automatically is to develop an algorithm that can automate the TP adjustment of a KTS. The manual adjustment process is converted into an MDP from the grid operators' perspective, and a model-free method is proposed to determine the optimal adjustment strategy. The proposed method uses the target TP ranges of all concerned KTSs as its input and outputs the adjusted power flow result. Unlike traditional model-based methods, the proposed method requires less manual interventions. In this paper, we only consider the active power adjustment.

The contributions of this paper are listed as follows:

- A Markov Decision Process is constructed from the operators' perspective to formulate the tie-line power adjustment problem based on a specific mapping strategy.
- A DRL based model-free method is proposed to generate the optimal generator adjustment strategy. The input information only contains the target KTSs and the related TP ranges.
- The ''stepwise training'' and the ''prioritized target replay'' are proposed to decompose large-scale complex problems and improve training efficiency.

The rest of this paper is presented as follows. The problem formulation for tie-line power adjustment is introduced in section II. Section III presents a new adaptive mapping strategy for tie-line power adjustment. Section IV proposes a model-free method based on DRL to learn the optimal adjustment strategy. In section V, experimental results demonstrate the effectiveness of the proposed method. In the end, section VI provides the concluding remarks.

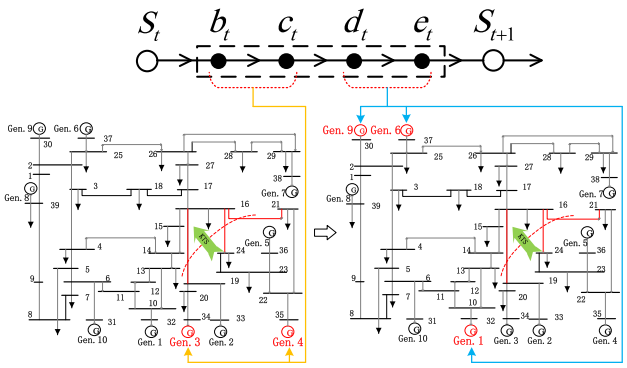

FIGURE 1. TP adjustment process in actual projects. S_t is the state of the power flow at time step t. b_t, c_t, d_t, and e_t are the actions executed between adjacent states.

II. PROBLEM FORMULATION

A. TIE-LINE POWER ADJUSTMENT PROCESS

A general tie-line power (TP) adjustment process is illustrated in Fig. 1. The S_t and S_{t+1} represent the power flow state at the

current time step t and the next time step $t+1$. In between S_t and S_{t+1}, there are commonly four actions as the generator choosing (b_t), generator state setting (c_t), generator group selecting (d_t) and power compensating (e_t). Fig. 1 takes the IEEE 39-bus system as an example. The three red tie-lines constitute one KTS, and the green arrow shows the positive direction. Gen 3 or Gen 4 is changed in b_t and c_t. Gen 1, Gen 6, and Gen 9 are under adjustment of d_t and e_t.

After the power system state is transformed from S_t to S_{t+1}, the actions b_{t+1}, c_{t+1}, d_{t+1}, and e_{t+1} will be executed until the TP of the target key transmission section (KTS) reaches its target value. It is worth mentioning that power flow convergence should be guaranteed at each time step.

B. MDP FORMULATION

The Markov Decision Process (MDP) is a mathematical structure for decision-making process modeling. It can be denoted as a five-tuple (S, A, P, R, γ), where S is the system state, A is the action set, P is the state transition probability, R is the immediate reward, and γ is a discount factor [28].

The TP adjustment problem is a decision-making process. Inspired by [21], we formulated the TP adjustment as a finite MDP. Given a target tie-line power $\mathrm{P}_C^{m,\mathrm{tar}}$ of the KTS m, the adjustment action is then determined in every time step. For instance, at time step t, the system state s_t, which contains the current power flow state and the adjustment target is observed. Based on s_t, an adjustment action a_t is executed. Then, a new system state s_{t+1} is obtained, and action a_{t+1} is executed until the target TP is achieved. Details of the MDP components and formulation for the TP adjustment are shown as follows.

1) System State: The system state at time step t is defined as a vector $s_t = (m, \mathrm{P}^t_{G,\tau_1}, \mathrm{P}^t_{G,\tau_2}, \ldots, \mathrm{P}^t_{G,\tau_{N_G}}, \mathrm{P}_C^{m,\mathrm{tar}})$, $\tau_i \in \Omega$. The vector s_t contains three types of variables: (1) m indicates the sequence number of a target KTS (a label to distinguish different KTSs); (2) P^t_{G,τ_i} represents the adjustable generators' injected power at time step t; Ω is the bus set to the adjustable generators (not for the slack bus); (3) $\mathrm{P}_C^{m,tar}$ denotes the target tie-line power of the KTS m.

2) Action: For the system state s_t, the adjustments (b_t, c_t, d_t, and e_t in Fig. 1) on the adjustable generators are represented by a real number a_t within $[a_{\min}, a_{\max}]$. The action a_t is executed according to a mapping strategy which is detailed in section III.

3) State Transition: The state transition is shown as

$$s_{t+1} = f(s_t, a_t), \tag{1}$$

where the state transition is determined by a_t and the corresponding mapping strategy together.

4) Reward: Due to the diversity of different tasks, the design of reward mechanisms is so difficult that there has not been a general principle for reward designing [29], [30].

In the TP adjustment problem, it is vital to make sure that the power flow calculation is convergent, and the output power of the slack bus generator is within its rated value. The TP adjustment should be executed based on these preconditions. Therefore, we define the reward function as below:

$$r(s_t, a_t) = \begin{cases} r_{\max}, & \text{cond.1 \& cond.2 \& cond.3} \\ -\left|\mathrm{P}_C^{m,tar} - \mathrm{P}_C^{m,t}\right|, & \text{cond.2 \& cond.3} \\ r_{\min}, & \text{others} \end{cases} \tag{2}$$

where cond.1 represents the tie-line power of the KTS m is achieved within $\left[\mathrm{P}_C^{m,tar} - \delta, \mathrm{P}_C^{m,tar} + \delta\right]$ (δ is the error range); cond.2 represents the output power of the slack bus generator is within its available range; cond.3 denotes the power flow calculation is convergent. Moreover, $r_{\max}$ should be non-negative, and $r_{\min}$ should be smaller than $\min\left(-\left|\mathrm{P}_C^{m,tar} - \mathrm{P}_C^{m,t}\right|\right)$. If cond.1, cond.2, and cond.3 are satisfied simultaneously, it means not only the target TP is achieved, but also the power flow state is reasonable. The maximum non-negative reward $r_{\max}$ is given. If only cond.2 and cond.3 are satisfied, it means the power flow state is reasonable, but the target TP is not achieved. A negative reward $-\left|\mathrm{P}_C^{m,tar} - \mathrm{P}_C^{m,t}\right|$ relating to the difference between $\mathrm{P}_C^{m,t}$ and $\mathrm{P}_C^{m,tar}$ is given as a punishment. If the power flow state is not reasonable (others), the minimum negative reward $r_{\min}$ is given as the most severe punishment.

5) Actor Function: The action a_t is determined by the adjustment strategy function μ shown as

$$a_t = \mu(s_t), \tag{3}$$

where s_t represents the system state as introduced above. $\mu(s_t)$ is called the *actor function* mapping s_t to a specific action a_t.

6) Critic Function: At time step t, the effect of action a_t is estimated by the sum of its expected future reward shown as

$$Q^{\mu}(s_t, a_t) = E_{\mu}\left[\sum_{k=0}^{\infty} \gamma^k \cdot r_{t+k} \,\middle|\, s_t, a_t\right], \tag{4}$$

where $Q^{\mu}(s_t, a_t)$ is called the *critic function*; $0 \le \gamma \le 1$ is the discount factor balancing the immediate and the future reward. Particularly, when $\gamma = 1$, the future reward is considered as important as the immediate reward. When $\gamma = 0$, only the immediate reward is considered [21].

The objective of the MDP formulation is to find the optimal actor function $a_t = \mu^*(s_t)$, based on which the adjustment strategy can always earn the largest expected reward estimated by the critic function.

III. MAPPING STRATEGY

In practice, operators adjust the high sensitive generators and try to change the power flow state in a minimum action. The general TP adjustments (b_t, c_t, d_t, and e_t) between adjacent states are actions in different decision spaces. To formulate the MDP for TP adjustments, a new mapping strategy is proposed to transform the sequential actions b_t, c_t, d_t, and e_t into one action a_t with a fixed decision space. Since the pre-learned human knowledge benefits the training efficiency

a lot [31], [32], we apply a similar idea by integrating the operators' work experience in the mapping strategy. Then, the pre-learned mapping strategy is divided into two parts: data preparation and dynamic mapping. Data preparation helps find the sensitive and insensitive generators to lower the searching difficulty (similar to b_t and d_t). Dynamic mapping helps execute the specific adjustments on generators (similar to c_t and e_t).

A. DATA PREPARATION

1) SENSITIVITY INDEX

As shown in (5), six sensitivity-related indexes are proposed to screen out generators:

$$\Delta \mathrm{P}_{C,i}^{m,pos} = \max(\mathrm{P}_{C,i}^{m,\max} - \mathrm{P}_C^m, \mathrm{P}_{C,i}^{m,\min} - \mathrm{P}_C^m), m \in \Omega_C, i \in \Omega \tag{5a}$$

$$\Delta \mathrm{P}_{C,i}^{m,neg} = \max(\mathrm{P}_C^m - \mathrm{P}_{C,i}^{m,\max}, \mathrm{P}_C^m - \mathrm{P}_{C,i}^{m,\min}), m \in \Omega_C, i \in \Omega \tag{5b}$$

$$\Delta \mathrm{P}_{C,i}^{m,ban} = \left|\Delta \mathrm{P}_{C,i}^{m,pos}\right| + \left|\Delta \mathrm{P}_{C,i}^{m,neg}\right|, \quad m \in \Omega_C, i \in \Omega \tag{5c}$$

where m represents the sequence number of a KTS; Ω_C is the set of the KTSs; P_C^m denotes the initial TP; $\mathrm{P}_{C,i}^{m,\max}$ ($\mathrm{P}_{C,i}^{m,\min}$) represents the TP of the KTS m by only setting the generator i to its maximum (minimum) value (In this paper, "generator i" represents the generator at bus i). $\Delta \mathrm{P}_{C,i}^{m,pos}$ ($\Delta \mathrm{P}_{C,i}^{m,neg}$) is the remaining adjustable amount of the KTS m in the positive (negative) direction resulting from generator i. $\Delta \mathrm{P}_{C,i}^{m,ban}$ represents the overall contribution of generator i.

$$S_{C,i}^{m,pos} = \begin{cases} \left|\dfrac{\Delta \mathrm{P}_{C,i}^{m,pos}}{\mathrm{P}_{G,i}^{\max} - \mathrm{P}_{G,i}}\right|, \mathrm{P}_{C,i}^{m,\max} \geq \mathrm{P}_{C,i}^{m,\min}, \mathrm{P}_{G,i}^{\max} \neq \mathrm{P}_{G,i} \\ \left|\dfrac{\Delta \mathrm{P}_{C,i}^{m,pos}}{\mathrm{P}_{G,i}^{\min} - \mathrm{P}_{G,i}}\right|, \mathrm{P}_{C,i}^{m,\max} < \mathrm{P}_{C,i}^{m,\min}, \mathrm{P}_{G,i}^{\min} \neq \mathrm{P}_{G,i} \\ 0, \text{others} \end{cases} \tag{5d}$$

$$S_{C,i}^{m,neg} = \begin{cases} \left|\dfrac{\Delta \mathrm{P}_{C,i}^{m,neg}}{\mathrm{P}_{G,i}^{\max} - \mathrm{P}_{G,i}}\right|, \mathrm{P}_{C,i}^{m,\max} \geq \mathrm{P}_{C,i}^{m,\min}, \mathrm{P}_{G,i}^{\max} \neq \mathrm{P}_{G,i} \\ \left|\dfrac{\Delta \mathrm{P}_{C,i}^{m,neg}}{\mathrm{P}_{G,i}^{\min} - \mathrm{P}_{G,i}}\right|, \mathrm{P}_{C,i}^{m,\max} < \mathrm{P}_{C,i}^{m,\min}, \mathrm{P}_{G,i}^{\min} \neq \mathrm{P}_{G,i} \\ 0, \text{others} \end{cases} \tag{5e}$$

$$S_{C,i}^{m,ban} = S_{C,i}^{m,pos} + S_{C,i}^{m,neg} \tag{5f}$$

where $\mathrm{P}_{G,i}$ is the initial power of generator i; $\mathrm{P}_{G,i}^{\max}$ ($\mathrm{P}_{G,i}^{\min}$) denotes the generator's maximum (minimum) power; $\mathrm{S}_{C,i}^{m,pos}$ ($\mathrm{S}_{C,i}^{m,neg}$) is a sensitivity index of generator i in the positive (negative) direction; $\mathrm{S}_{C,i}^{m,ban}$ denotes the overall index.

2) GENERATOR RANKING

Based on (5), the adjustable generators rank in three sequences for each KTS, as shown in (6):

$$\Psi_{pos}^m = \{\alpha_k | \Delta \mathrm{P}_{C,\alpha_k}^{m,pos} > \Delta \mathrm{P}_{C,\alpha_{k+1}}^{m,pos} \text{ or } (\Delta \mathrm{P}_{C,\alpha_k}^{m,pos} = \Delta \mathrm{P}_{C,\alpha_{k+1}}^{m,pos} \& \mathrm{S}_{C,\alpha_k}^{m,pos} \geq \mathrm{S}_{C,\alpha_{k+1}}^{m,pos}), \alpha_k \in \Omega, k = 1, 2, 3, \ldots, N_G, \} \tag{6a}$$

$$\Psi_{neg}^m = \{\beta_k | \Delta \mathrm{P}_{C,\beta_k}^{m,neg} > \Delta \mathrm{P}_{C,\beta_{k+1}}^{m,neg} \text{ or } (\Delta \mathrm{P}_{C,\beta_k}^{m,neg} = \Delta \mathrm{P}_{C,\beta_{k+1}}^{m,neg} \& \mathrm{S}_{C,\beta_k}^{m,neg} \geq \mathrm{S}_{C,\beta_{k+1}}^{m,neg}), \beta_k \in \Omega, k = 1, 2, 3, \ldots, N_G, \} \tag{6b}$$

$$\Psi_{ban}^m = \{\gamma_k | \Delta \mathrm{P}_{C,\gamma_k}^{m,ban} < \Delta \mathrm{P}_{C,\gamma_{k+1}}^{m,ban} \text{ or } (\Delta \mathrm{P}_{C,\gamma_k}^{m,ban} = \Delta \mathrm{P}_{C,\gamma_{k+1}}^{m,ban} \& \mathrm{S}_{C,\gamma_k}^{m,ban} \leq \mathrm{S}_{C,\gamma_{k+1}}^{m,ban}), \gamma_k \in \Omega, k = 1, 2, 3, \ldots, N_G, \} \tag{6c}$$

where N_G denotes the number of the adjustable generators; Ψ_{pos}^m, Ψ_{neg}^m, and Ψ_{ban}^m are sets of generators ranked depending on the proposed sensitivity indexes. Specifically, Ψ_{pos}^m (Ψ_{neg}^m) is for the adjustment in the positive (negative) direction, and Ψ_{ban}^m for the power compensation. To improve the astringency of power flow calculation, generators that lead to non-convergence while calculating $\mathrm{P}_{C,i}^{m,\max}$ or $\mathrm{P}_{C,i}^{m,\min}$ should be removed. In this paper, the subscripts α_k, β_k, and γ_k only denote the sorted generators from Ψ_{pos}^m, Ψ_{neg}^m, and Ψ_{ban}^m respectively.

3) GENERATORS FOR POWER COMPENSATION

The subset ψ_{ban}^m is prepared for power compensation (d_t and e_t). The possible maximum and minimum boundaries of the power fluctuation resulting from the TP adjustments (b_t and c_t) are calculated according to (7).

$$n_{pos}^m = \min N, \quad s.t. \sum_{k=1}^{N} \Delta \mathrm{P}_{C,\alpha_k}^{m,pos} \geq \varepsilon_c \cdot (\mathrm{P}_C^{m,\max} - \mathrm{P}_C^m), \alpha_k \in \Psi_{pos}^m \tag{7a}$$

$$n_{neg}^m = \min N, \quad s.t. \sum_{k=1}^{N} \Delta \mathrm{P}_{C,\beta_k}^{m,neg} \geq \varepsilon_c \cdot (\mathrm{P}_C^m - \mathrm{P}_C^{m,\min}), \beta_k \in \Psi_{neg}^m \tag{7b}$$

$$Z_{G,i}^m = \begin{cases} 1, & \mathrm{P}_{C,i}^{m,\max} \geq \mathrm{P}_{C,i}^{m,\min} \\ -1, & \mathrm{P}_{C,i}^{m,\max} < \mathrm{P}_{C,i}^{m,\min} \end{cases} \tag{7c}$$

$$\Delta \mathrm{P}_{G,sum}^{m,\max} = \max\left(\sum_{k=1}^{n_{pos}^m} \left(\mathrm{P}_{G,\alpha_k}^{\max} - \mathrm{P}_{G,\alpha_k}\right) \cdot \frac{1 + Z_{G,\alpha_k}^m}{2}, \sum_{k=1}^{n_{neg}^m} \left(\mathrm{P}_{G,\beta_k}^{\max} - \mathrm{P}_{G,\beta_k}\right) \cdot \frac{1 - Z_{G,\beta_k}^m}{2}\right) \tag{7d}$$

$$\Delta \mathrm{P}_{G,sum}^{m,\min} = \max\left(\sum_{k=1}^{n_{pos}^{m}}\left(\mathrm{P}_{G,\alpha_k} - \mathrm{P}_{G,\alpha_k}^{\min}\right)\cdot\frac{1-Z_{G,\alpha_k}^{m}}{2},\ \sum_{k=1}^{n_{neg}^{m}}\left(\mathrm{P}_{G,\beta_k} - \mathrm{P}_{G,\beta_k}^{\min}\right)\cdot\frac{1+Z_{G,\beta_k}^{m}}{2}\right) \tag{7e}$$

where $\mathrm{P}_{C}^{m,\max}$ ($\mathrm{P}_{C}^{m,\min}$) is the maximum (minimum) value of the target TP of the KTS $m \in \Omega_C$; n_{pos}^{m} (n_{neg}^{m}) denotes the number of generators required in b_t and c_t; ε_c is a reliable coefficient to guarantee enough generators in ψ_{ban}^{m} ($\varepsilon_c \geq 1$, a higher ε_c means more candidate generators); $Z_{G,i}^{m}$ is a relationship index. $\Delta \mathrm{P}_{G,\mathrm{sum}}^{m,\max}$ and $\Delta \mathrm{P}_{G,\mathrm{sum}}^{m,\min}$ are the boundaries of the injected power fluctuation resulting from b_t and c_t.

Then, ψ_{ban}^{m} is optimized by (8) to obtain the sets of generators that affect the tie-line power of the KTS m as little as possible.

Object function:

$$\min \sum_{i=1}^{n^{com}} \Delta \mathrm{P}_{C,\gamma_{x_i}}^{m,ban} + \sum_{i=1}^{n^{up}} \Delta \mathrm{P}_{C,\gamma_{y_i}}^{m,ban} + \sum_{i=1}^{n^{down}} \Delta \mathrm{P}_{C,\gamma_{z_i}}^{m,ban} \tag{8a}$$

Subjected to :

$$\sum_{i=1}^{n^{com}}\left(\mathrm{P}_{G,\gamma_{x_i}}^{\max} - \mathrm{P}_{G,\gamma_{x_i}}\right) + \sum_{i=1}^{n^{up}}\left(\mathrm{P}_{G,\gamma_{y_i}}^{\max} - \mathrm{P}_{G,\gamma_{y_i}}\right) \geq \Delta \mathrm{P}_{G,sum}^{m,\min} \tag{8b}$$

$$\sum_{i=1}^{n^{com}}\left(\mathrm{P}_{G,\gamma_{x_i}} - \mathrm{P}_{G,\gamma_{x_i}}^{\min}\right) + \sum_{i=1}^{n^{down}}\left(\mathrm{P}_{G,\gamma_{z_i}} - \mathrm{P}_{G,\gamma_{z_i}}^{\min}\right) \geq \Delta \mathrm{P}_{G,sum}^{m,\max} \tag{8c}$$

$$\mathrm{P}_{G,\gamma_{x_i}}^{\max} \geq P^{thr},\ \mathrm{P}_{G,\gamma_{y_i}}^{\max} - \mathrm{P}_{G,\gamma_{y_i}} \geq P^{thr},\ \mathrm{P}_{G,\gamma_{z_i}} - \mathrm{P}_{G,\gamma_{z_i}}^{\min} \geq P^{thr} \tag{8d}$$

$$\begin{cases} 1 \leq x_i < x_{i+1} \leq n^{com},\ n^{com} \in Z \\ n^{com} \leq y_i < y_{i+1} \leq n^{up},\ n^{up} \in \mathrm{Z} \\ n^{com} \leq z_i < z_{i+1} \leq n^{down},\ n^{down} \in Z \\ \psi_{ban}^{m} = \{\gamma_{x_i}\} \cup \{\gamma_{y_i}\} \cup \{\gamma_{z_i}\}, \\ \qquad \gamma_{x_i}, \gamma_{y_i}, \gamma_{z_i} \in \Psi_{ban}^{m} \end{cases} \tag{8e}$$

and (5a) (5b) (5c) (8f)

where P^{thr} is a threshold value for the remaining adjustment amount; n^{com}, n^{up}, n^{down}, γ_{x_i}, γ_{y_i} and γ_{z_i} are the unknown variables to be solved; ψ_{ban}^{m} is a subset of Ψ_{ban}^{m}, and the elements of ψ_{ban}^{m} can be denoted as

$$\psi_{ban}^{m} = \left\{\gamma_{k_i} \middle| k_i < k_{i+1}, i \in Z^{+}, k_i \in \{x_i\} \cup \{y_i\} \cup \{z_i\}, \gamma_{k_i} \in \Psi_{ban}^{m}\right\} \tag{9}$$

To decrease the calculation complexity, ψ_{ban}^{m} can also be optimized by using the heuristic method that selects generators γ_{k_i} according to (8d) from $i = 1$ until (8b) and (8c) are met.

B. DYNAMIC MAPPING

The dynamic mapping (DM) process consists of the active mapping (AM) and the passive mapping (PM). The active mapping includes the actions of b_t and c_t. The passive mapping contains the actions of d_t and e_t.

1) ACTIVE MAPPING

The subsets (ψ_{pos}^{m} and ψ_{neg}^{m}) of sensitive generators are dynamically formulated based on $\mathrm{P}_{C}^{m,tar}$ from Ψ_{pos}^{m} and Ψ_{neg}^{m} to improve the training efficiency according to (10).

$$\psi_{pos}^{m} = \{\alpha_1, \ldots, \alpha_k, \ldots, \alpha_{n_{pos}^{m,tar}} | \alpha_k \in \Psi_{pos}^{m}, n_{pos}^{m,tar} = \min N, \ \mathrm{s.t.} \sum_{k=1}^{N} \Delta \mathrm{P}_{C,\alpha_k}^{m,pos} \geq \varepsilon_c \cdot \left|\mathrm{P}_{C}^{m,tar} - \mathrm{P}_{C}^{m}\right|, \mathrm{P}_{C}^{m,tar} \geq \mathrm{P}_{C}^{m}, k \in Z^{+}\} \tag{10a}$$

$$\psi_{neg}^{m} = \{\beta_1, \ldots, \beta_k, \ldots, \beta_{n_{neg}^{m,tar}} | \beta_k \in \Psi_{neg}^{m}, n_{neg}^{m,tar} = \min N, \ \mathrm{s.t.} \sum_{k=1}^{N} \Delta \mathrm{P}_{C,\beta_k}^{m,neg} \geq \varepsilon_c \cdot \left|\mathrm{P}_{C}^{m,tar} - \mathrm{P}_{C}^{m}\right|, \mathrm{P}_{C}^{m,tar} < \mathrm{P}_{C}^{m}, k \in Z^{+}\} \tag{10b}$$

where ε_c is the same with (7a) and (7b).

It is worth mentioning that ψ_{pos}^{m} and ψ_{neg}^{m} are composed of the generators with higher sensitivity to the KTS m, and ψ_{ban}^{m} consists of the generators with lower sensitivity. In practice, ψ_{pos}^{m} and ψ_{neg}^{m} usually contain a small number of generators and will not include the same generators with ψ_{ban}^{m}. If ψ_{pos}^{m} or ψ_{neg}^{m} contains common elements with ψ_{ban}^{m}, it means the target TP range of the KTS m is too broad and is not available for the power system.

As shown in (11), the action space is dynamically divided by a^{i} depending on $\mathrm{P}_{C}^{m,tar}$. a^{1} is always set to $a^{\min}$.

$$a^{i+1} = \begin{cases} a^{\min} + \dfrac{\sum_{j=1}^{i} \Delta \mathrm{P}_{C,\alpha_j}^{m,pos}}{\sum_{j=1}^{n_{pos}^{m,tar}} \Delta \mathrm{P}_{C,\alpha_j}^{m,pos}}\left(a^{\max} - a^{\min}\right), \mathrm{P}_{C}^{m,tar} \geq \mathrm{P}_{C}^{m}, i = 1, 2, \ldots, n_{pos}^{m,tar} \\ a^{\min} + \dfrac{\sum_{j=1}^{i} \Delta \mathrm{P}_{C,\alpha_j}^{m,neg}}{\sum_{j=1}^{n_{neg}^{m,tar}} \Delta \mathrm{P}_{C,\alpha_j}^{m,neg}}\left(a^{\max} - a^{\min}\right), \mathrm{P}_{C}^{m,tar} < \mathrm{P}_{C}^{m}, i = 1, 2, \ldots, n_{neg}^{m,\mathrm{tar}} \end{cases} \tag{11}$$

If $P_C^{m,tar} \geq P_C^m$, the active mapping result is calculated according to (12a), or else (12b).

$$P_{G,\alpha_i}^t = \begin{cases} P_{G,\alpha_i} + \dfrac{a_t - a^i}{a^{i+1} - a^i}\left(P_{G,\alpha_i}^{\max} - P_{G,\alpha_i}\right), P_{C,\alpha_i}^{\max} \geq P_{C,\alpha_i}^{\min}, \\ \qquad a_t \in \left[a^i, a^{i+1}\right) \\ P_{G,\alpha_i} + \dfrac{a_t - a^i}{a^{i+1} - a^i}\left(P_{G,\alpha_i}^{\min} - P_{G,\alpha_i}\right), P_{C,\alpha_i}^{\max} < P_{C,\alpha_i}^{\min}, \\ \qquad a_t \in \left[a^i, a^{i+1}\right) \end{cases} \tag{12a}$$

$$P_{G,\beta_i}^t = \begin{cases} P_{G,\beta_i} + \dfrac{a_t - a^i}{a^{i+1} - a^i}\left(P_{G,\beta_i}^{\min} - P_{G,\beta_i}\right), P_{C,\beta_i}^{\max} \geq P_{C,\beta_i}^{\min}, \\ \qquad a_t \in \left[a^i, a^{i+1}\right) \\ P_{G,\beta_i} + \dfrac{a_t - a^i}{a^{i+1} - a^i}\left(P_{G,\beta_i}^{\max} - P_{G,\beta_i}\right), P_{C,\beta_i}^{\max} < P_{C,\beta_i}^{\min}, \\ \qquad a_t \in \left[a^i, a^{i+1}\right) \end{cases} \tag{12b}$$

where $\alpha_i \in \psi_{pos}^m$ ($\beta_i \in \psi_{neg}^m$) represents the selected generator; P_{G,α_i}^t (P_{G,β_i}^t) denotes the output power of generator α_i (β_i) under action a_t.

2) PASSIVE MAPPING

The passive mapping compensates for the power fluctuation and helps alleviate the problem of non-convergence and unreasonable operation state. At time step $t+1$, if the output power of generator i is larger than time step t, the passive mapping is executed according to (13a), or else (13b).

$$\begin{aligned} P_{G,i}^{t+1} - P_{G,i}^t &= P_{G,\gamma_{k_{n+1}}}^t - P_{G,\gamma_{k_{n+1}}}^{t+1} + \sum_{j=1}^{n}\left(P_{G,\gamma_{k_j}}^t - P_{G,\gamma_{k_j}}^{\min}\right), \\ P_{G,i}^{t+1} &\geq P_{G,i}^t, P_{G,\gamma_{k_{n+1}}}^{t+1} \leq P_{G,\gamma_{k_{n+1}}}^t, \gamma_{k_j} \in \psi_{ban}^m \end{aligned} \tag{13a}$$

$$\begin{aligned} P_{G,i}^t - P_{G,i}^{t+1} &= P_{G,\gamma_{k_{n+1}}}^{t+1} - P_{G,\gamma_{k_{n+1}}}^t + \sum_{j=1}^{n}\left(P_{G,\gamma_{k_j}}^{\max} - P_{G,\gamma_{k_j}}^t\right), \\ P_{G,i}^{t+1} &< P_{G,i}^t, P_{G,\gamma_{k_{n+1}}}^{t+1} \geq P_{G,\gamma_{k_{n+1}}}^t, \gamma_{k_j} \in \psi_{ban}^m \end{aligned} \tag{13b}$$

where i denotes the active mapping generator; $P_{G,i}^t$ is the output power of generator i at time step t. In (13a), to balance the increasing injected power, the first n generators in ψ_{ban}^m are set to their minimum power and the $(n+1)th$ generator γ_{n+1} is set to $P_{G,\gamma_{n+1}}^{t+1}$. Analogously, in (13b), to balance the decreasing injected power, the first n generators in ψ_{ban}^m are set to their maximum power and the $(n+1)th$ generator γ_{n+1} is set to $P_{G,\gamma_{n+1}}^{t+1}$. Except for the first $n+1$ generators, all the others in ψ_{ban}^m remain unchanged.

Fig. 2 shows the whole process of the mapping strategy. The data preparation is executed only once before the TP adjustment. However, the dynamic mapping is repetitively activated based on a_t and $P_C^{m,tar}$, in which the passive mapping depends on the result from the active mapping.

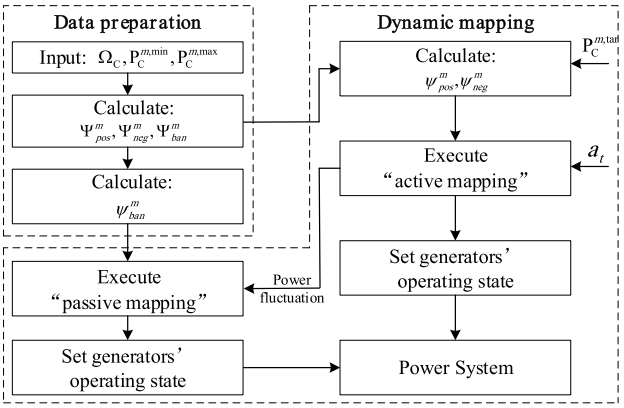

FIGURE 2. Logic block diagram of the mapping strategy.

The efficiency improvement of the mapping strategy is discussed in Case Study I.

IV. THE PROPOSED DRL FOR MDP

To solve the formulated MDP, a model-free method is proposed based on an actor-critic structure [28]. The critic function is iteratively updated based on (14), and the actor function is updated under the critic function [28].

$$Q^{\mu}(s_t, a_t) = E_{\mu}\left[r(s_t, a_t) + \gamma Q^{\mu}(s_{t+1}, \mu(s_{t+1}))\right] \tag{14}$$

Since the action space is continuous, and the power flow features are high-dimensional, The deep neural networks (DNNs) are utilized to approximate the actor and critic functions. The overall diagram is presented in Fig. 3. The input information is the system state s_t and fed into the actor network for the action a_t generation. Experience tuples (s_t, a_t, r_t, s_{t+1}) are stored in the experience replay buffer. Based on these tuples, the actor, the critic, and the target networks are updated in a sequence. Finally, the optimal adjustment strategy is given by the trained actor network.

A. THE DEEP NEURAL NETWORK ARCHITECTURE

1)Actor network: The actor network $\mu(s_t|\theta^{\mu})$ is a multi-layer fully-connected neural network which can extract the discriminative features from the input system state vector s_t and output action a_t. The input layer is fully connected to the first hidden layer with V_1 units,

$$v_1 = g\left(W_0 \cdot s_t^T + b_0\right), \tag{15a}$$

where W_0 is the matrix of weights; T represents the transposition operation; b_0 is the biases; g denotes the rectified linear activation function [33],

$$g(x) = \min(\max(x, 0), 6) \tag{15b}$$

Similarly, the hidden layers are also fully-connected with each other based on (15c),

$$v_{i+1} = g(W_i \cdot v_i + b_i), \quad i = 1, 2, \ldots, n-1, \tag{15c}$$

where W_i and b_i are the weights and biases of the ith hidden layer; n is the number of the hidden layers, and v_i is the value

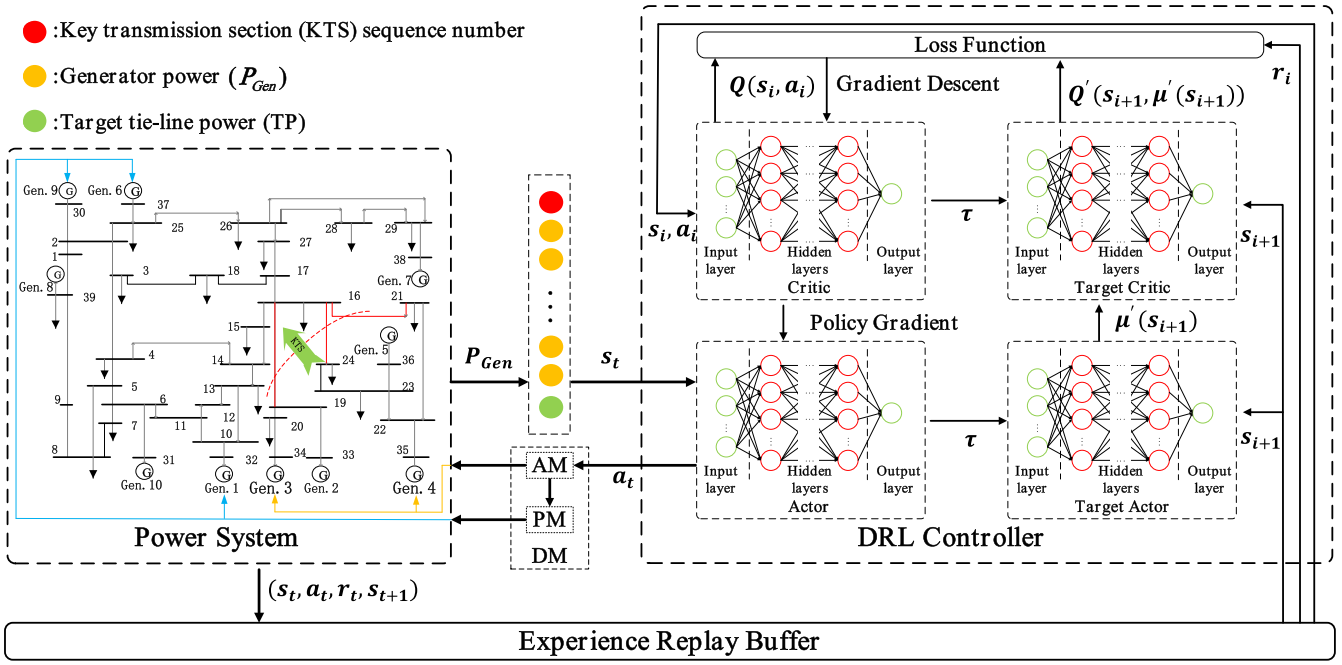

FIGURE 3. The overall diagram of the proposed method for TP adjustment. The actor network generates adjustment actions. The critic network estimates the quality of the current actor network. The target actor network and the target critic network improve training stability. DM represents the dynamic mapping. AM denotes the active mapping. PM is the passive mapping. The 39-bus system is taken as an example. Gen. 3 and Gen. 4 are for the active mapping. Gen. 1, Gen. 6, and Gen. 9 are for the passive mapping.

of the ith hidden units. The output of the actor network is action a_t.

$$a_t = f\left(W_n \cdot v_n + b_n\right) \tag{15d}$$

where W_n is the weights, b_n is the biases, and f is the hyperbolic tangent function. Therefore, action a_t can be bounded within a finite range easily.

2)Critic network: The critic network denoted by $Q^{\mu}\left(s_t, a_t | \theta^{Q}\right)$ is also a multi-layer fully-connected neural network. Its input information is a combination of s_t and a_t, and the value of its first hidden unit is

$$v_1 = g\left(W_0 \cdot (s_t, a_t)^T + b_0\right). \tag{16a}$$

The hidden layers of the critic network share the same structure with the actor network.

As mentioned above, the critic network is to approximate the expected reward of the current adjustment strategy, and its output value cannot be bounded. So its output layer is fully connected to the last hidden layer with no activation function, as shown in (16b).

$$Q^{\mu}\left(s_t, a_t\right) = W_n \cdot v_n + b_n \tag{16b}$$

3)Target networks: Since the critic network being updated is also used to estimate the expected reward, the updating process is prone to divergence. Similar to the target network in [16], we create the target actor and critic networks. They are copies of the actor and critic networks and denoted by $\mu'\left(s_t | \theta^{\mu'}\right)$ and $Q'\left(s_t, a_t | \theta^{Q'}\right)$ respectively.

Instead of directly copying the weights, the target networks are updated by tracking the learned networks $\theta' \leftarrow \tau\theta + (1-\tau)\theta'$ with $\tau \ll 1$. In this way, the target networks change slowly, and the stability of updating is improved significantly [17].

B. TRAINING OF THE DEEP NEURAL NETWORKS

Algorithm 1 shows how to train the proposed DNNs to adjust the tie-line power of the KTSs automatically. The parameters of the actor network, the critic network, the target actor network, and the target critic network are denoted as θ^{μ}, θ^{Q}, $\theta^{\mu'}$, and $\theta^{Q'}$ respectively. The inputs of Algorithm 1 contain the information of the target KTSs and the target TP ranges ($P_C^{m,\max}$ and $P_C^{m,\min}$). Its outputs are the parameters of the four proposed DNNs.

In line 1 of Algorithm 1, the data for training is prepared according to Section III-A. Then, in line 2 and 3, the four proposed DNNs are initialized. In line 4, an empty experience replay buffer is initialized as $\mathcal{H}$ with its size being N. When the replay buffer is full, the oldest samples will be discarded [17]. Besides, P_1 is set to 1 as the initial priority for the first transition tuple [36]. In line 5, a correlated noise is initialized with the Ornstein-Uhlenbeck process [34] for the exploration policy [35]. After that, the parameters θ^{μ}, θ^{Q}, $\theta^{\mu'}$, and $\theta^{Q'}$ are updated in turns until the proposed test is passed, which is to verify whether the trained DNNs can achieve all the given objectives within a preset accuracy (the outer loop starting from line 6). Each episode begins at time step 1 and ends at time step $T_{\max}$ or somewhere between

1 and T_{max} when the reward r_t equals r_{max}. At the beginning of each episode, as shown in line 7, the initial system state s_t is obtained (containing a random KTS sequence number and a random $P_C^{m,tar}$). Then, in the inner loop starting from line 8, the TP adjustment strategy is scheduled step by step. At each time step, the action $a_t = \mu(s_t|\theta^\mu)$ is selected based on the ε-greedy search method [28], i.e., noise $\mathcal{N}$ (Ornstein-Uhlenbeck process) is added to action a_t with probability ε whose initial value is set to 1. Then, the probability ε is updated by multiplying a coefficient σ (line 9). It is worth mentioning that $a'_t = a_t + \mathcal{N}$ should also be bounded within $[a_{min}, a_{max}]$, and the minimum value of ε is set to 0.1. In line 10, action a_t is executed according to the dynamic mapping, and then the reward r_t and the new state s_{t+1} is obtained.

After that, the transition (s_t, a_t, r_t, s_{t+1}) is stored in the replay buffer $\mathcal{H}$ and labeled with the current maximal priority p_t (line 11)

$$p_t = \max(p_i), \quad p_i \in \mathcal{H}, \tag{17}$$

where $p_i > 0$ represents the priority of transition i in $\mathcal{H}$, and its initial value P_1 is set to 1 [36]. Then, a minibatch of K transitions are sampled from $\mathcal{H}$ based on probability (line 12). Readers can refer to [36] for more details about the sampling method and the calculation of ω_j and δ_j.

After the four DNNs are updated (line 13 to 15) [17], the current reward $r(s_t, a_t)$ is checked, and the current episode will terminate when $r_t = r_{max}$. Finally, the proposed test for the discrete TP values of each KTS is executed every M episodes (line 20). In this paper, we assume the testing TP value increases linearly from $P_C^{m,min}$ to $P_C^{m,max}$ with a fixed step size ΔM. The parameters θ^μ, θ^Q, $\theta^{\mu'}$, and $\theta^{Q'}$ will be outputted when the test is passed.

It is worth noting that using the experience replay buffer can not only contribute to better data efficiency by sampling the transitions multiple times but also improve the stability of the training process by breaking the temporal correlation between the transitions [16]. Moreover, the priority-sampling process (line 12) can improve the efficiency of DNN updating as well.

C. EXECUTING TIE-LINE POWER ADJUSTMENT

The TP adjustment process is presented in Algorithm 2. The system state s_t contains the information of the power flow state and the adjustment target. Its output is the final power flow state.

D. PERFORMANCE IMPROVEMENT TECHNIQUES

1) STEPWISE TRAINING

As in Algorithm 1, all target TPs can be trained together as all information included in the input. However, as the number of the KTSs or the range of the TP is enormous in a large-scale power system, the efficiency of training will decline dramatically due to the limit fitting ability of a specific DNN. Or even worse, the proposed DNNs are not able to learn the whole adjustment strategy.

Algorithm 1 Training of the Deep Neural Networks

Input: Target KTS, target TP range $P_C^{m,max}$ and $P_C^{m,min}$.
Output: DNNs' parameters θ^μ, θ^Q, $\theta^{\mu'}$, $\theta^{Q'}$

1: Data preparation.
2: Randomly initialize the actor network θ^μ and the critic network θ^Q.
3: Initialize the target actor and critic networks with weights $\theta^{\mu'} \leftarrow \theta^\mu$ and $\theta^{Q'} \leftarrow \theta^Q$.
4: Initialize the experience replay buffer $\mathcal{H} = \emptyset, p_1 = 1$.
5: Initialize an Ornstein-Uhlenbeck process.
6: **while** fail to pass the proposed test **do**
7: Obtain the initial system state.
8: **for** $t = 1{:}T_{max}$ **do**
9: Select action a_t based on ε-greedy search and update ε.
10: Execute a_t based on the dynamic mapping, observe reward $r(s_t, a_t)$ and process to the new state s_{t+1}.
11: Store transition (s_t, a_t, r_t, s_{t+1}) in $\mathcal{H}$ with maximal priority $p_t = max(p_i), p_i \in \mathcal{H}$.
12: Sample a minibatch of K transitions from $\mathcal{H}$ based on priorities and update the priority-sampling parameters.
13: Update the critic network by minimizing the accumulative loss $L(\theta^Q) = \sum_{j=1}^{K} \delta_j^2 \cdot \omega_j / K$
14: Update the actor network using the sampled policy gradient:

$$\nabla_{\theta^\mu} J \approx \frac{1}{K} \sum_{j=1}^{K} \nabla_{\mu(s_j)} Q(s_j, \mu(s_j)|\theta^Q) \cdot \nabla_{\theta^\mu} \mu(s_j|\theta^\mu)$$

15: Update weights of the two target networks.

$$\theta^{Q'} \leftarrow \tau\theta^Q + (1-\tau)\theta^{Q'}$$
$$\theta^{\mu'} \leftarrow \tau\theta^\mu + (1-\tau)\theta^{\mu'}$$

16: **if** $r(s_t, a_t) = r_{max}$
17: **break**
18: **end if**
19: **end for**
20: Execute the proposed test every M episodes.
21: **end while**
22: **Return:** θ^μ, θ^Q, $\theta^{\mu'}$, $\theta^{Q'}$.

To avoid the abovementioned situation of DNN training or redesign of more complex DNNs, "stepwise training" is proposed as an improved technique for large-scale power systems by dividing the whole training targets into several parts. That is dividing the target TP range of each KTS into smaller subintervals and then apply Algorithm 1, respectively.

The proposed DNNs for each subinterval share the same structures, and their weights are different due to independent training. Nonetheless, the size of each subinterval also affects the training substantially. From any initial power flow state,

Algorithm 2 Execute Tie-Line Power Adjustment

Input: Initial power flow state, target KTS sequence number, and the target tie-line power $P_C^{m,tar}$.
Output: The achieved power flow state.
1: Load the actor network's parameters θ^μ (trained by Algorithm 1).
2: **for** Time step $t = 1$:T_{max} **do**
3: Obtain the system state s_t.
4: Feed state s_t into the actor network and calculate the adjustment action $a_t = \mu(s_t|\theta^\mu)$.
5: Execute the dynamic mapping based on a_t..
6: **if** the target TP $P_C^{m,tar}$ is achieved
7: **break**
8: **end if**
9: **end for**
10: **Return:** the adjusted power flow state.

the subintervals should be exactly covered by adjusting a certain number R_g of generators in ψ_{pos}^m or ψ_{neg}^m, where R_g is a positive integer.

2) PRIORITIZED TARGET REPLAY

Inspired by the idea of [36], the ''prioritized target replay'' is proposed to improve training efficiency.

As shown in Algorithm 1, the target KTS and the target TP $P_C^{m,tar}$ are selected randomly. Since the target range of each KTS is fixed, it will help improve the training efficiency by replaying the tie-line power that fails to pass the proposed test with an appropriate probability ε'.

The proposed technique defines the tie-line power failed to be achieved as the prioritized target, which can be dynamically obtained every M episodes from the proposed test. Besides, to improve the actor network's generalization ability, random noise should be added to the prioritized target.

$$P_C^{m,tar,p} = P_C^{m,tar,t} + \lambda \tag{18}$$

In (18), $P_C^{m,tar,t}$ is the prioritized target. λ denotes the noise in mean distribution within $[\lambda_{min}, \lambda_{max}]$. $P_C^{m,tar,p}$ represents the noisy prioritized target to be replayed.

V. EXPERIMENTAL RESULTS

In this section, the proposed method is demonstrated on the IEEE 39-bus system [37] and an actual power grid in a certain area of China. The general simulation setups are presented in Section V-A. Section V-B shows four case studies. Finally, the performance of the ''prioritized target replay'' is analyzed in Section V-C.

A. GENERAL EXPERIMENTAL SETUP

Two key transmission sections (KTSs) are first tested in the 39-bus system. As shown in Fig. 5, the KTS 1 (consisting of tie-lines (16,19), (16,21), and (16,24)) tie-line power (TP) range is set to [200MW, 1400MW]. Similarly, the KTS 2 (consisting of tie-line (3,4)) TP range is set to [−200MW, 400MW]. All the generators' rated power is set to 1100MW. Besides, the KTS 1 and 2 initial TPs are 828MW and 37MW, respectively.

The actual power grid contains 2725 buses, 5 DC lines, 722 generators, and 979 loads. As shown in Fig. 8, the KTS 1 (including five AC lines in area JL) TP range is set to [0MW, 2800MW] and the KTS 2 (including five AC lines in area LN) TP range is set to [100MW, 2800MW]. The KTS 1 and 2 initial TPs are 1538MW and 1539MW, respectively.

The (target) actor network contains five layers (one input layer, three hidden layers, and one output layer). The dimensions of each layer are (12, 400, 600, 100, 1) for the 39-bus system and (322, 400, 600, 100, 1) for the actual 2725-bus system. Similarly, the (target) critic network also contains five layers. The dimensions of each layer are (13, 400, 600, 100, 1) for the 39-bus system and (323, 400, 600, 100, 1) for the actual 2725-bus system. The four DNNs are updated based on the Adam algorithm [38].

The minibatch K is set to 32 in the 39-bus system and 64 in the actual 2725-bus system. According to lots of experiments, though ε_c may be different for different power systems, the suitable ε_c for most situations usually remains between 1.1 and 1.3. In this paper, we set ε_c to 1.2 for both systems. Other relevant hyperparameters for both power systems are set as follows: $\sigma = 0.99999$, $N = 5000$, $\delta = 10$MW, $\gamma = 0.9$, $\tau = 0.00005$, $r_{max} = 100$, $r_{min} = -100$, $a_{max} = 1$, $a_{min} = -1$, $R_g = 1$, $\lambda_{min} = -5$MW, $\lambda_{max} = 5$MW, $\Delta M = 10$MW, and $M = 100$.

The numerical tests are performed on the computer with one 1080Ti GPU and one i8700K CPU. The code is written in Python with TensorFlow (an open source package). Pandapower [27] is utilized for the power flow calculation. After the training process, the proposed approach can be deployed for adjusting the tie-line power of each KTS.

B. EXPERIMENTAL RESULTS

1) CASE STUDY I

In this case study, we evaluate the proposed method on the 39-bus system and discuss the effect of using the dynamic mapping (DM). The adjustment strategy for the KTS 1 and 2 is trained together.

Firstly, the proposed DNNs are trained with DM applied for 45,100 episodes. Every 100 episodes a test is performed to check whether the actor network has been trained well enough. The solid lines in Fig. 4 show the evolution of the average cumulative reward (ACR) and the average adjustment error (AAE). The action is selected based on the ε-greedy search. In the training process, the probability ε declines from 1.0 to 0.1 gradually. As presented in Fig. 4, during the first 7,500 episodes, the ACR increases fast, and the AAE decreases dramatically. Then, from episode 7,500 to episode 20,000, the ACR fluctuates between 36.78 and 90.46. After episode 20,000, the ACR converges around 88 with small oscillations. Finally, in episode 45,100, the ACR converges to 99.99, and the AAE declines to 4MW.

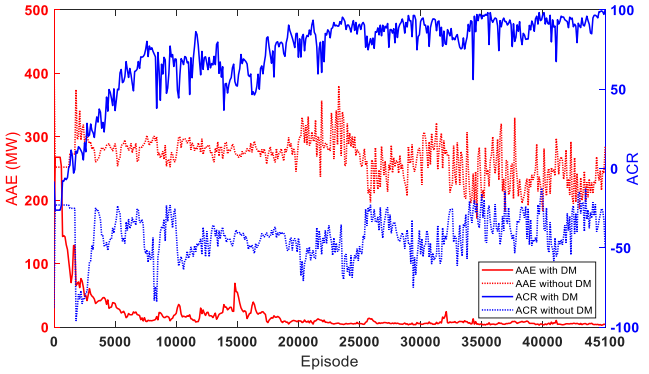

FIGURE 4. The IEEE 39-bus system training processes without applying improvement techniques.

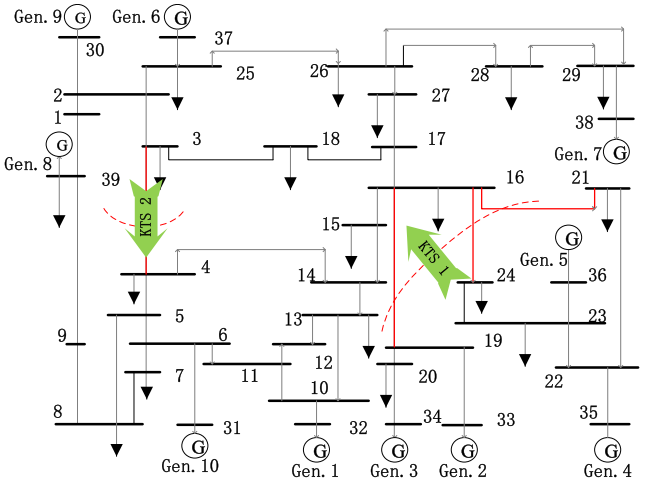

FIGURE 5. The KTS 1 and the KTS 2 in the IEEE 39-bus system. The green arrows show the positive directions.

Then, the proposed DNNs are trained again with DM removed, and the action space is divided equally for every adjustable generator. The dot lines in Fig. 4 show that the training without DM does not achieve better results. After 45,100 episodes, the ACR still remains below zero, and the AAE is around 230MW. It will be more time consuming to achieve a satisfying performance.

Comparing the presented plots, it demonstrates that the proposed method succeeds in learning an adjustment strategy to maximize the cumulative reward in the training process, and the dynamic mapping (DM) is of great importance to the training efficiency.

2) CASE STUDY II

In this case study, the proposed approach is evaluated on the 39-bus system with the improvement techniques applied. The probability ε' of "prioritized target replay" is set to 0.5. Different from Case Study I, the improved method automatically divides the whole training process into four parts and trains DNNs independently.

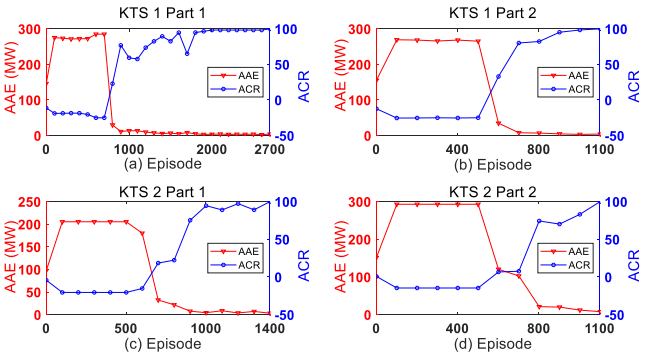

FIGURE 6. Training process on the IEEE 39-bus system with improvement techniques. (a) and (b) are for the KTS 1; (c) and (d) are for the KTS 2.

a: TRAINING PROCESS

Fig. 6 shows the details of the processes for each training part. In the first 500 episodes of each training part, the ACR is very low and flat, and the AAE is very high. Then, after episode 500, the ACR increases, and the AAE decreases rapidly. The training episodes of each part are 2700, 1100, 1400, and 1100 respectively, and the total episodes equal 6300, which is much less than Case Study I (45,100). This result demonstrates that the proposed method works in different subintervals, and the improvement techniques can improve the training efficiency significantly.

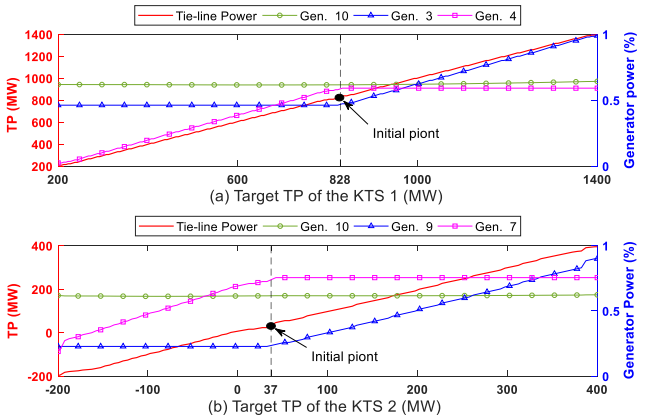

FIGURE 7. Verifications on the trained DNNs of the 39-bus system. (a) is for the KTS 1 and (b) is for the KTS 2. The "Initial point" indicates the initial tie-line power.

b: PERFORMANCE EVALUATION

The trained DNNs are verified on the whole target ranges for the rightness. Fig. 7 shows the test results, where the x-axis represents the target TP of each KTS; the y-axis on the left denotes the achieved TP of each KTS; the y-axis on the right is the percentage of the generator output power. The red line denotes the TP achieved from the trained DNNs, and the green line with circle marks represents Gen. 10, which is the generator at the slack bus (bus 31).

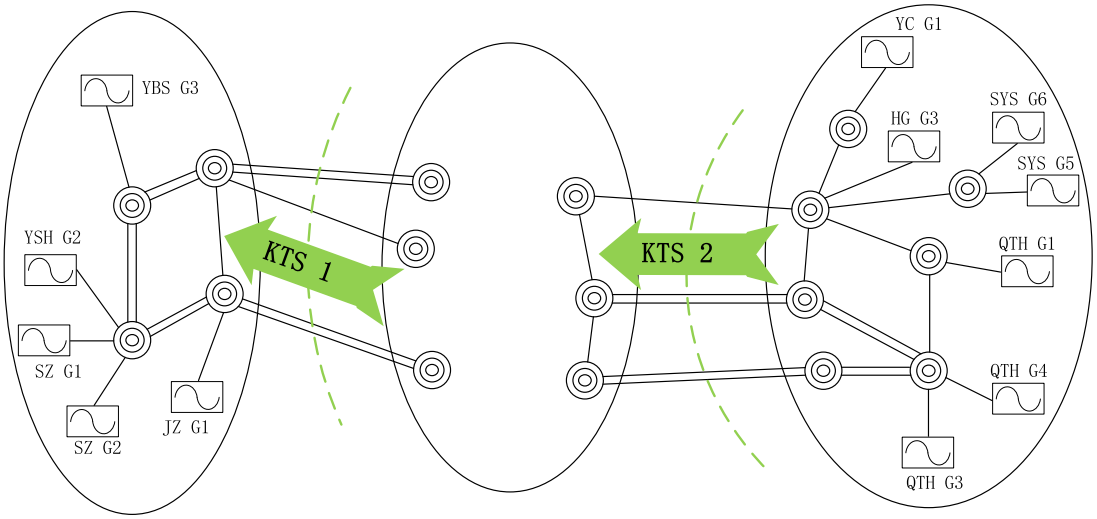

FIGURE 8. Simplified wiring diagram of the actual 2725-bus power grid. Both the KTS 1 and the KTS 2 consist of five tie-lines. The green arrows show the positive directions.

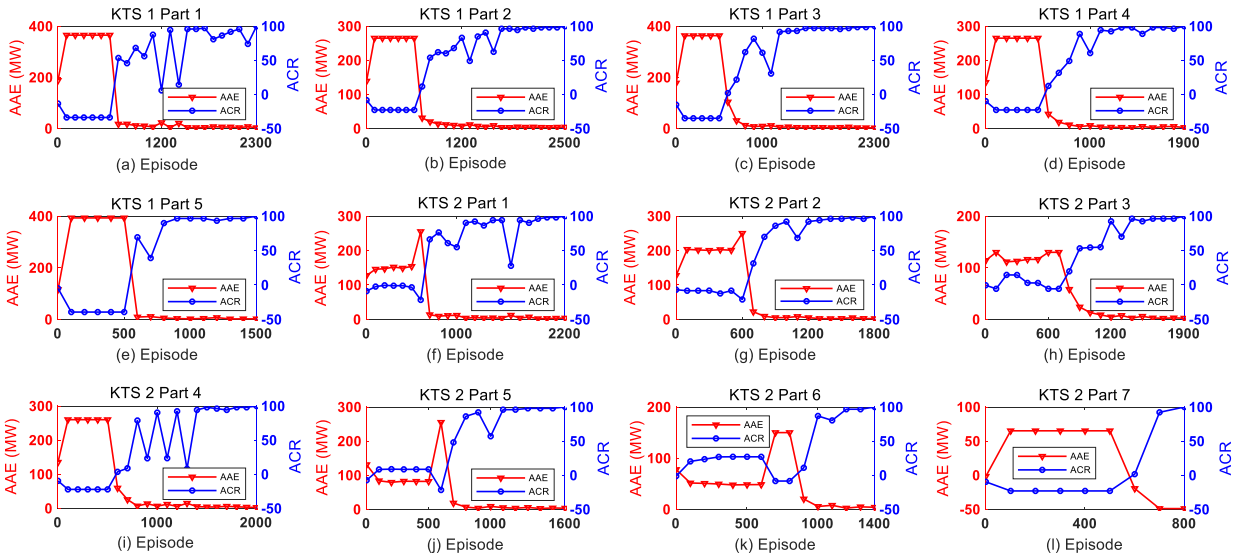

FIGURE 9. Training processes on the actual 2725-bus system with improvement techniques. (a) to (e) are for the KTS 1, and (f) to (l) are for the KTS 2.

In Fig. 7 (a), as the target TP increases to 1400MW, Gen. 3 is selected, and its output power increases from 508MW (46.2% rated power) to 1088.16 MW (98.9% rated power). As the target TP decreases to 200MW, Gen. 4 is selected, and its output power decreases from 650MW (59.1% rated power) to 26MW (2.4% rated power). Similarly, in Fig. 7 (b), as the target TP increases to 400MW, Gen. 9 is selected, and its output power increases from 250MW (22.7% rated power) to 989.5MW (90.9% rated power). As the target TP decreases to -200MW, Gen. 7 is selected, and its output power decreases from 830MW (75.5% rated power) to 207.5MW (18.9% rated power).

All the generators selected for each subinterval own the largest remaining adjustment amount for each specific target. Besides, the output power of Gen. 10 remains stable (around 62% rated power) in both Fig. 7 (a) and (b) due to passive mapping. This result demonstrates that the proposed method can continuously and flexibly adjust the tie-line power.

3) CASE STUDY III

In this case study, the proposed method is evaluated on the actual 2725-bus system with the improvement techniques applied. The probability ε' of "prioritized target replay" is also set to 0.5. Different from Case Study II, the whole training process is automatically divided into 12 parts, as shown in Fig. 9.

a: TRAINING PROCESS

Fig. 9 (a) to (e) are for the KTS 1, and (f) to (l) for the KTS 2. Similarly, the ACR of each training part remains very low and flat in around the first 500 episodes. Then, the

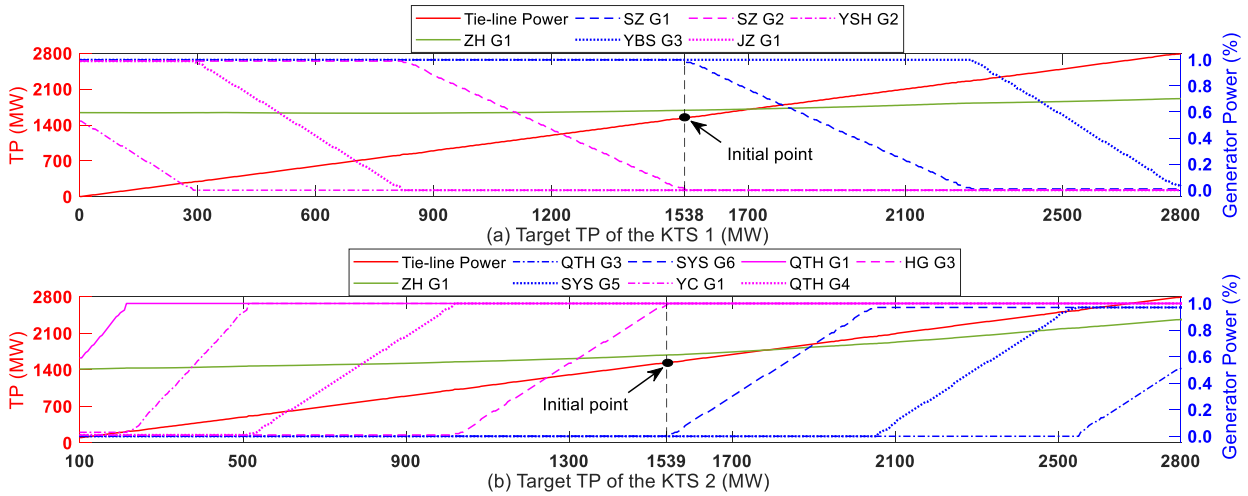

FIGURE 10. Verifications on the trained DNNs of the actual 2725-bus system. (a) is for the KTS 1, and (b) is for the KTS 2. The "initial point" indicates the initial tie-line power.

ACR begins to increase and reach close to 100 quickly, and the AAE decreases almost to 0 in the end. The number of the total training episodes is 22,200, which is about half of Case Study I, even with a larger TP searching range. This result demonstrates that the proposed method suits the actual power system.

b: PERFORMANCE EVALUATION

As shown in Fig. 10, the trained DNNs are tested on the actual 2725-bus system, and the meaning of the coordinate axes is the same as Fig. 7. The red line denotes the achieved TP, and the green line represents the slack bus generator (ZH G1). Table 1 shows supplementary information about the generators selected by the active mapping.

TABLE 1. Information of the generators selected by active mapping.

Generator	In service	Initial power (MW)	Max power (MW)
SZ G1	Yes	800	800
SZ G2	NO	0	800
YBS G3	YES	600	600
YSH G2	NO	0	600
JZ G1	NO	0	600
SYS G5	NO	0	600
SYS G6	NO	0	600
QTH G1	YES	350	350
QTH G3	NO	0	600
QTH G4	YES	600	600
HG G3	YES	600	600
YC G1	YES	350	350
ZH G1	YES	368	600

In Fig. 10 (a), as the target TP of the KTS 1 increases, two generators SZ G1 and YBS G3 (denoted by blue lines) are selected. From the initial tie-line power 1538MW to 2270MW, SZ G1 is chosen first, and its output power declines from its rated value 800MW to 8MW (1% rated power). From 2270MW to 2800MW, YBS G3 is selected, and its output power decreases from its rated value 600MW to 24MW (4% rated power). Then, the TP of the KTS 1 reaches 2784MW, and the adjustment error is only 16MW compared to 2800MW.

Contrastively, as the target TP of the KTS 1 decreases, three out-service generators SZ G2, JZ G1, and YSH G2 are selected, as shown in purple lines. As the tie-line power decreases from 1538MW to 0MW, the three generators are set in service one by one, and their ultimate power is set to 792MW, 594MW, and 318MW (99%, 99%, and 53% rated power), respectively. In the end, the final TP reaches 4.3MW, and the adjustment error is only 4.3MW. During the test process, the output power of ZH G1 (the slack bus generator) fluctuates between 354MW and 412MW (59% and 70.2% rated power) due to the passive mapping. The maximum adjustment error compared to the target value is 16MW.

Similarly, Fig. 10 (b) shows the test results of the KTS 2. In the increasing direction, SYS G6, SYS G5, and QTH G3 are set in service in turns. Then, the TP reaches 2793MW, and the adjustment error is only 7MW compared to 2800MW. In the decreasing direction, the output power of HG G3, QTH G4, YC G1, and QTH G1 are turned down successively. After that, the achieved TP is 103MW, and the adjustment error is only 3MW compared to 100MW. The output power of ZH G1 fluctuates between 303MW and 527MW (50.5% and 87.8% rated power) accordingly.

This case study demonstrates that the proposed method can adjust the tie-line power of a large-scale power system flexibly. Furthermore, the largest adjustment error of this test is within 16MW, which can meet the engineering requirement.

4) CASE STUDY IV

In this section, the proposed approach is compared with the interior point method (IPM) based on which the TP adjustment is formulated as an optimal power flow (OPF) problem. The standard OPF is inherited from MATPOWER 7.0 [39]. The TPs of the KTSs are set as the extra constraints based

on DC and AC network model (denoted by IPM-DC and IPM-AC), respectively.

The model-based method is tested on the KTS 1 of the 39-bus system and the KTS 1 of the actual 2725-bus system. To simulate the actual situation, a specific group of generators are selected beforehand for adjustment based on operators' experience, i.e. six generators (Gen. 1, Gen. 3, Gen. 4, Gen. 6, Gen. 9, and Gen. 10) for the 39-bus system, and 22 generators for the actual 2725-bus system. All the other generators remain unadjusted.

TABLE 2. KTS 1 of the IEEE 39-bus system.

Target TP (MW)	TP adjustment error (MW)		
	DRL	IPM-DC	IPM-AC
200	7.9	42.1	0
400	-4.4	51.1	0
600	2.2	69.5	0
800	1.8	-8.9	0
1000	3.3	-53.9	0
1200	7.1	-60.9	0

TABLE 3. KTS 2 of the actual 2725-bus system.

Target TP (MW)	TP adjustment error (MW)		
	DRL	IPM-DC	IPM-AC
0	4.3	180.6	N/A
400	-2.3	261.4	0
800	0.9	345.9	N/A
1600	-3.5	315.1	0
2000	5.5	159.3	0
2700	-3.1	182.6	N/A

The comparison results are shown in Table 2 and 3, respectively. Notice that the adjustment error of IPM-DC is larger than the proposed method (DRL), especially in the actual 2725-bus system. As shown in Table 3, when the tie-line power is set to 0MW, 800MW, and 2700MW, the model-based IPM-AC does not converge. Besides, the model-based method depends on the operators' experience to pre-determine a group of candidate generators for adjustment. In contrast, the proposed method (DRL) can achieve the TP adjustment with better accuracy and less pre-determined work.

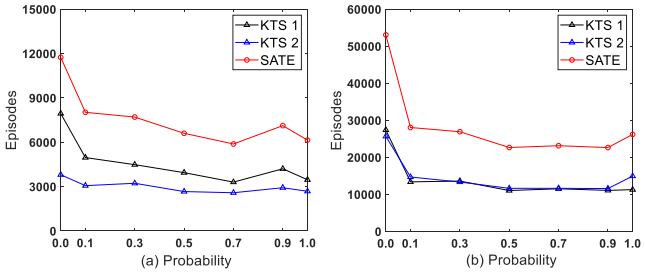

FIGURE 11. Effects of the "prioritized target replay." (a) is for the IEEE 39-bus system and (b) is for the actual 2725-bus system.

C. FURTHER EXPERIMENT

In this section, further experiments are presented about the probability ε' of "prioritized target replay." Case Study II and III are examined with different values of ε', and each ε' is tested five times. Fig. 11 shows the average total training episodes: the black line for the KTS 1, the blue line for the KTS 2, and the red line for the sum of the average total training episodes (SATE) of each KTS. Besides, more results are shown in Table 4. From Fig. 11 and Table 4, it is easy to conclude that the probability ε' of "prioritized target replay" can affect the training efficiency significantly, and the more complex the training task is, the higher the efficiency will be improved. According to the experimental results, it is better to set ε' no less than 0.3.

TABLE 4. Information of average training episodes.

Probability	Average training episodes	
	IEEE-39 bus	Actual power grid
0.0	11740	53040
0.1	8020	28040
0.3	7700	26900
0.5	6600	22640
0.7	5880	23120
0.9	7120	22620
1.0	6140	26180

VI. CONCLUSION

In this paper, a mapping strategy is proposed from the operators' perspective to formulate the tie-line power adjustment problem as a Markov Decision Process (MDP) with unknown transition probability. Then, a model-free method based on deep reinforcement learning (DRL) is introduced to determine the optimal adjustment strategy. The presented method uses an actor-critic structure with the "stepwise training" and the "prioritized target replay" to decompose training scale and improve the training efficiency. Experimental results demonstrate that the presented method is capable of learning and adjusting the KTS tie-line power with only the target range information. Furthermore, the comparison to a traditional model-based approach demonstrates the higher accuracy and better adaptability of the proposed method.

In the future plan, the mapping strategy will be explored further together with the reward function design, the reactive power constraints, and more limitations of the adjustable generators to refine the TP adjustment process.

REFERENCES

[1] B. Obama, "The irreversible momentum of clean energy," *Science*, vol. 355, no. 6321, pp. 126–129, 2017.

[2] E. Du, N. Zhang, B.-M. Hodge, Q. Wang, C. Kang, B. Kroposki, and Q. Xia, "The role of concentrating solar power toward high renewable energy penetrated power systems," *IEEE Trans. Power Syst.*, vol. 33, no. 6, pp. 6630–6641, Nov. 2018.

[3] J. Yan, H. Zhang, Y. Liu, S. Han, L. Li, and Z. Lu, "Forecasting the high penetration of wind power on multiple scales using multi-to-multi mapping," *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 3276–3284, May 2018.

[4] X. Fang, B.-M. Hodge, E. Du, C. Kang, and F. Li, "Introducing uncertainty components in locational marginal prices for pricing wind power and load uncertainties," *IEEE Trans. Power Syst.*, vol. 34, no. 3, pp. 2013–2024, May 2019.

[5] Y. Zhang, X. Han, B. Xu, M. Wang, P. Ye, and Y. Pei, "Risk-based admissibility analysis of wind power integration into power system with energy storage system," *IEEE Access*, vol. 6, pp. 57400–57413, 2018.

[6] Y. Wang, N. Zhang, C. Kang, M. Miao, R. Shi, and Q. Xia, "An efficient approach to power system uncertainty analysis with high-dimensional dependencies," *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 2984–2994, May 2018.

[7] F. Tiutiunnyk, A. Prystupa, and V. Bodunov, "Improving methods for evaluating the stability of electrical systems with distributed generation," in *Proc. 2nd Int. Young Sci. Forum Appl. Phys. Eng. (YSF)*, Kharkiv, Ukraine, Oct. 2016, pp. 37–40.

[8] Y. Zhang, B. Liu, B. Luo, C. Cheng, and J. Yang, "PSD-BPA based automatic integration and adjustment method of power grid operation plan data," (in Chinese), *Autom. Electr. Power Syst.*, vol. 41, no. 1, pp. 102–108, Jan. 2017.

[9] W. Hongfu, M. Shixia, W. Yi, and Z. Zhiqiang, "An approximate power flow method to deal with the non-convergence problem of power flow calculation," in *Proc. Int. Conf. Power Syst. Technol.*, Guangzhou, China, Nov. 2018, pp. 292–299.

[10] L. Ren and P. Zhang, "Generalized microgrid power flow," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 3911–3913, Jul. 2018.

[11] S. S. Reddy and J. A. Momoh, "Realistic and transparent optimum scheduling strategy for hybrid power system," *IEEE Trans. Smart Grid*, vol. 6, no. 6, pp. 3114–3125, Nov. 2015.

[12] J.-O. Lee, Y.-S. Kim, E.-S. Kim, and S.-I. Moon, "Generation adjustment method based on bus-dependent participation factor," *IEEE Trans. Power Syst.*, vol. 33, no. 2, pp. 1959–1969, Mar. 2018.

[13] S. Wang and S. Gao, "Available transfer capability analysis method of AC–DC power system based on security region," *J. Eng.*, vol. 2019, no. 16, pp. 2386–2390, Mar. 2019.

[14] Y. Liu, J. Zhao, L. Xu, T. Liu, G. Qiu, and J. Liu, "Online TTC estimation using nonparametric analytics considering wind power integration," *IEEE Trans. Power Syst.*, vol. 34, no. 1, pp. 494–505, Jan. 2019.

[15] S. Xu and S. Miao, "Calculation of TTC for multi-area power systems based on improved ward-PV equivalents," *IET Gener., Transmiss. Distrib.*, vol. 11, no. 4, pp. 987–994, Mar. 2017.

[16] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, pp. 529–533, Feb. 2015.

[17] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2016, pp. 1–14.

[18] R. Yousefian, R. Bhattarai, and S. Kamalasadan, "Transient stability enhancement of power grid with integrated wide area control of wind farms and synchronous generators," *IEEE Trans. Power Syst.*, vol. 32, no. 6, pp. 4818–4831, Nov. 2017.

[19] J. Zhang, C. Lu, J. Si, J. Song, and Y. Su, "Deep reinforcement learning for short-term voltage control by dynamic load shedding in China Southem power grid," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Rio de Janeiro, Brazil, Jul. 2018, pp. 1–8.

[20] Z. Yan and Y. Xu, "Data-driven load frequency control for stochastic power systems: A deep reinforcement learning method with continuous action search," *IEEE Trans. Power Syst.*, vol. 34, no. 2, pp. 1653–1656, Mar. 2019.

[21] Z. Wan, H. Li, H. He, and D. Prokhorov, "Model-free real-time EV charging scheduling based on deep reinforcement learning," *IEEE Trans. Smart Grid*, vol. 10, no. 5, pp. 5246–5257, Sep. 2019.

[22] F. Ruelens, B. J. Claessens, S. Quaiyum, B. De Schutter, R. Babuška, and R. Belmans, "Reinforcement learning applied to an electric water heater: From theory to practice," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 3792–3800, Jul. 2018.

[23] R. Lu, S. H. Hong, and M. Yu, "Demand response for home energy management using reinforcement learning and artificial neural network," *IEEE Trans. Smart Grid*, vol. 10, no. 6, pp. 6629–6639, Nov. 2019. doi: 10.1109/TSG.2019.2909266.

[24] E. Mocanu, D. C. Mocanu, P. H. Nguyen, A. Liotta, M. E. Webber, M. Gibescu, and J. G. Slootweg, "On-line building energy optimization using deep reinforcement learning," *IEEE Trans. Smart Grid*, vol. 10, no. 4, pp. 3698–3708, Jul. 2019.

[25] B. Claessens, P. Vrancx, and F. Ruelens, "Convolutional neural networks for automatic state-time feature extraction in reinforcement learning applied to residential load control," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 3259–3269, Jul. 2018.

[26] D. Zhang, X. Han, and C. Deng, "Review on the research and practice of deep learning and reinforcement learning in smart grids," *CSEE J. Power Energy Syst.*, vol. 4, no. 3, pp. 362–370, Sep. 2018.

[27] *Pandapower 2.0*. Accessed: Apr. 15, 2019. [Online]. Available: http://www.pandapower.org/

[28] R. S. Sutton and A. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 2018.

[29] D. Dewey, "Reinforcement learning and the reward engineering principle," in *Proc. AAAI Spring Symp.*, Stanford, CA, USA, 2014, pp. 1–8.

[30] D. Hadfield-Menell, S. Milli, P. Abbeel, S. J. Russell, and A. Dragan, "Inverse reward design," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 6765–6774.

[31] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis, "Mastering the game of Go with deep neural networks and tree search," *Nature*, vol. 529, no. 7587, pp. 484–489, Jan. 2016.

[32] Q. Fettes, M. Clark, R. Bunescu, A. Karanth, and A. Louri, "Dynamic voltage and frequency scaling in NoCs with supervised and reinforcement learning techniques," *IEEE Trans. Comput.*, vol. 68, no. 3, pp. 375–389, Mar. 2019.

[33] A. Krizhevsky and G. Hinton, "Convolutional deep belief networks on cifar-10," unpublished, 2010, vol. 40, no. 7, pp. 1–9.

[34] G. E. Uhlenbeck and L. S. Ornstein, "On the theory of the Brownian motion," *Phys. Rev.*, vol. 36, no. 5, p. 823, Sep. 1930.

[35] P. Wawrzyński, "Control policy with autocorrelated noise in reinforcement learning for robotics," *Int. J. Mach. Learn. Comput.*, vol. 5, no. 2, pp. 91–95, 2015.

[36] T. Schaul, J. Quan, I. Antonoglou, and D. Silver, "Prioritized experience replay," 2015, *arXiv:1511.05952*. [Online]. Available: https://arxiv.org/abs/1511.05952

[37] *IEEE 39-Bus Test Case Archive*. Accessed: Dec. 17, 2018. [Online]. Available: https://icseg.iti.illinois.edu/ieee-39-bus-system/

[38] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: https://arxiv.org/abs/1412.6980

[39] *Matpower 7.0*. Accessed: Jul. 12, 2019. [Online]. Available: https://matpower.org/

HUATING XU received the B.S. degree in electrical engineering and automation from Sichuan University, Chengdu, China, in 2012. He is currently pursuing the M.S. degree with the China Electric Power Research Institute, Beijing, China. He was an Electrical Engineer with the China General Nuclear Power Group, from 2012 to 2017. His current research interests include deep reinforcement learning, reactive power and tie-line power adjustment, optimal power flow, and multiagent systems.

ZHIHONG YU received the Ph.D. degree in electrical engineering from the Harbin Institute of Technology, China, in 2004. She is currently with the China Electric Power Research Institute (CEPRI). She is also a Principal Engineer with the Dynamic Security Assessment Studies Group. Her current research interests include power system stability simulation, analysis, and control and data mining and its engineering applications in power systems.

QINGPING ZHENG received the B.S. degree in communication engineering from the Beijing University of Posts and Telecommunications, Beijing, China, in 2018, where he is currently pursuing the M.S. degree in electronic and communication engineering. His research interests include deep learning, deep reinforcement learning, and deep reinforcement learning applications in power systems.

JINXIU HOU received the B.S. degree in electrical and electronics engineering from Shanghai Electric Power University, China, in 2014, and the M.S. degree in power systems from the China Electric Power Research Institute, China (CEPRI), in 2017. He was a Power Systems Engineer with the CEPRI, from 2017 to 2019. He is currently an Applications Engineer with the CEPRI, where he is involved in the development of various projects and implementation, and the development of various DSA applications function development.

YAWEI WEI received the B.S. degree from the Shanghai University of Electric Power, Shanghai, China, in 2010, the M.S. degree from Michigan Technological University, Houghton, MI, USA, in 2014, and the Ph.D. degree from the Department of Electrical and Computer Engineering (ECE), Clemson University, SC, USA, in 2018, focusing on situational intelligence for improving power system operations under high penetration of photovoltaics.

ZHIJIAN ZHANG received the M.S. degree from North China Electric Power University. She is currently a Senior Engineer with the State Grid Beijing Electric Power Dispatching and Control Center, Beijing, China. She is mainly devoted to power grid dispatching and control technology.